METHOD OF ENCODING AND DECODING DATA AND DEVICE FOR REALIZATION OF THE METHOD

BACKGROUND OF THE INVENTION

The invention relates to means for protecting data from an unauthorized access, and may be used in crypto-systems for encoding, transferring and decoding communications, and in other systems for protection of data.

The prior art discloses engineering solutions providing protection of transmitted data by means of a special equipment or encoding software, for example, by using scrambler for protection of telephone conversations [1] pp. 35-37, Fig. 22. The scrambler operates on the principle on inversion of an audio signal. As a result of an inversion, a usual speech turns to a senseless gang of sounds, but the initial signal is accepted by the user without any distortion. The telephone set is equipped with the block for voice modification controlled by the encoder. The encoder stores 13122 user's codes providing 52488 digital combinations. The read-only memory of the set stores the resident software, which codes and decodes the transmitted data in several variants and controls the work of the whole set.

However, this prior art solution has problems in providing a fair degree of secrecy, since for disclosing the confidential codes it is enough to execute a limited number of mathematical operations that are fast and effectively executed by the modern high-speed electronic engineering.

The main characteristic of a crypto-system is the degree of secrecy. The task of a cryptographer is to provide the utmost secrecy and authenticity of the transferred data. Alternatively, a crypto-analyst "forces open", or "breaks", the crypto-system designed by a cryptographer. The crypto-analyst tries to decipher the set of encoded symbols and to deliver the encrypted communication as the plaintext.

Prior art discloses technical solutions for protecting the transferred data by using a specific device and/or encoding software. Known codes are based on two simple methods: substitution and interchange. Interchange uses simple mixing of plain-text symbols, the key of an interchange encryptor defines the specific type of mixing. The frequency distribution of individual symbols in the encoded text is identical to that of the plaintext. For substitution, each symbol of the plaintext is replaced by another symbol of the same alphabet, and the specific type of substitution is determined by the secret key.

For example, the algorithm in the Data Encryption Standard (DES) [2], p. 33-34 uses the both methods. The algorithm comprises plaintext, unencrypted text and the key as binary sequences having the length 64, 64 and 56 bits, respectively. When DES is used in an electronic book or table mode, the 64-bit blocks of the plaintext are encoded independently by using one key. The algorithm of DES includes 16 rounds or cycles, each of which has simple interchanges combined with substitution in four-bit groups. In each pass, 48 key bits are selected in a pseudo-random manner from the full 56-bit key.

The problem of DES is that this prior art solution does not provide a fair degree of secrecy, since for disclosure of such secret codes with possible number of 2^{64} keys combinations, substituting of all keys during a brute-force attack using modern computer techniques is performed in an acceptable time. Also, using the same plaintext and not varying the keys, produces the same encoded text. Analysis reveals the statistical regularity of the correlation between the plaintext and the encoded text, and may allow decoding the encoded text without using direct substitution of all the keys.

A crypto-system using public keys RSA is described in [2] p. 37-39. This system uses a one-way function - discrete logarithms raising to a power.

GOST P. 34.11 - 94 [3], p. 3-8 discloses hatching consisting in comparing an optional set of data as a sequence of binary symbols, with a short, fixed length image thereof. In this system 64-bit subwords are encoded using keys of 256 bit length.

The drawbacks of these systems are the small key length, which may permit decoding during acceptable time, and a slow decoding speed. These systems are practically stable systems.

Theoretically stable systems have perfect secrecy. According to Shannon [4] p. 333-402, that means that the plaintext, and the encoded text or cryptogram, are statistically independent for all plaintext and cryptograms.

A prior art Vernan crypto-system is a theoretically stable crypto-system. Theoretically stable systems make certain demands on a key. For a system with closed keys the indeterminacy of the key should not be less than the indeterminacy of the plaintext. In theoretically stable systems, the length of a key should be not less than the length of the plaintext. In the Vernan system, the key length is equal to the length of the plaintext. This system was used in a codenotebook [5] for transfer of one encoded text. This is the main drawback of a codebook because the key should be changed and delivered with every transfer.

There are known crypto-systems using the so-called randomisers [2] p. 26 - 27. A randomiser is a software or a hardware device that encodes some symbols of plaintext with some random plurality of codes.

Typically, this is done for providing equal frequency of the plaintext alphabet. Symbol frequency equalisation is required so that a crypto-analyst cannot organise decoding of a plaintext based on analysis of frequency characteristics of a cryptogram. For a random plaintext and a random selection of a code, a derandomiser should correctly determine the initial symbol without transfer of data from the randomiser location. In classical systems with a small randomising field, this task is solved by substituting codes belonging to the randomised symbol. Randomisers, however, do not play a substantial role in crypto-protectability of an encoding system, as secret keys are the main means of protection.

Under the combination of the essential features, the closest prior art object to the claimed method and device is disclosed in [6], which provides a device and method of encoding that uses a principle of full randomizing symbols of the initial alphabet on a plurality of codes with potencies of large dimensionality. This prior art invention was selected by the inventors for the prototype of the claimed invention.

Regarding the method for the selected prototype object, a method of encoding and transferring data, wherein the addressee is provided a key to the received communications with data on regularities corresponding to the values of the communication transmitted to him beforehand, with specific values of the initial data for the whole set of symbols of the said kind of data. The method further includes processing data using the said regularities and transferring, to the addressee, the communication containing data, obtained during processing data, the values of transmitted data, which depend on random generated numbers being calculated during processing data. The method is further characterized in that the addressee is provided with a set of functions $Y_1 \ldots Y_n = Y_i(X)$ beforehand, where X is a variable, and each Y_i corresponds to a specific symbol of data. The addressee is also provided with the support function $U = U(Z)$, where Z is a variable, and with the key function $W = W(Y, U)$, where Y and U are variables accepting values of any of the values from the values of the said functions Y_i and U. In the course of processing of a transmitted data for each symbol, there are generated two random numbers X and Z, the respective value of Y is calculated on basis of the respective function $Y_i(X)$ for a specific symbol. The value of U is further calculated on basis of the support function U (Z). The value of W for this symbol is calculated on basis of the key function W (Y, U) and obtained for the symbol value of Y and the value of U from the support function, and the addressee is transmitted the communication containing data on the thus obtained values of W, X and Z for each symbol of the initial data.

With regard to a device, the object selected for the prototype is a device for realizing a method of encoding and transferring data, which comprises a unit for data input, a set of symbols, a database on plurality of characteristic functions that transform the specific symbols with the communication. The database is supplied with a calculator connected to a generator of random numbers. The device further comprises an encoder and

a unit for recording and transmitting communications. The encoder is connected to the set of symbols and calculator output. The device further comprises a unit for calculating the values of the support function and a unit for calculating the values of the key function. The generator of random numbers is supplied with two outputs joint with the encoder. The first output of the generator of random numbers is connected also to the input of the unit for calculating values of the support function, and the second output is connected to the input of the calculator of the database on regularities, the regularities. The output of this calculator is connected to the encoder through the unit for calculating values of the key function, and the second input of the latter is connected to the output of the unit for calculating values of the support function.

However, the problem of the object selected for the prototype is that in the course of the encryption the length of the encrypted communication exceeds the length of the initial communication by several times.

SUMMARY OF THE INVENTION

The aim of the claimed invention is providing an improved method of encrypting by means of obtaining several communications from one initial, at least one of the obtained communications may be compressed up to preset sizes so that any connection between the initial text and the cryptogram is completely lost for a cryptoanalyst.

As a result of the solution to the problem, there is achieved a new technical effect consisting of creating a new system of encrypting that ensures a high cryptostability system without any increase in the length of the communication. The said technical effect is achieved as follows.

A method of encrypting data comprises the following steps:

- Preliminary generation of data on plurality of characteristic functions that transform the values of symbols of the initial communication with the specific values of the encrypted communication for the total set of values of the said kind of communications;

- determination the number (n) of transformation cycles of the initial communication;
- realization of the transformation cycle comprising:
- generation of the feature (R_i), defining regularity used for transformation of the communication in the current transformation cycle;
- transformation of the communication with use of the selected regularity;
- repetition of transformation cycles the certain number of times;
- transformation of the communication in each cycle being realized in a way resulting in forming cycle data (C_i), transformed in the said cycle and the accessory data (F_i) for the said cycle;
- the number (n) of transformation cycles of the initial communication is selected from the preset criterion;
- forming the encrypted communication consisting of two parts, one of which contains the finally transformed data (C_n), and second one contains accessory data array ($F = \{F_1, F_2, \ldots, F_n\}$) .

A further improvement of the method is characterized by:

- transformation of the communication in each cycle is realized in a way resulting in forming cycle data (C_i) transformed in the said cycle, being of the shorter or equal length with the initial communication, and resulting in forming an accessory data (F_i) for the said cycle;
- the number (n) of transformation cycles of the initial communication is selected from the preset criterion (for example, the size of the finally transformed data);
- forming the encrypted communication consisting of two parts, one of which contains the finally transformed data (C_n) being of shorter length than the initial communication, and second one contains accessory data array ($F = \{F_1, F_2, \ldots, F_n\}$).

Still a further improvement of the method is characterized by:

- transformation of the communication in each cycle realized in a way resulting in forming cycle data (C_i) transformed in the said cycle, being of shorter, equal or longer length with the initial communication, and resulting in forming an accessory data (F_i) for the said cycle;

- the number (n) of transformation cycles of the initial communication is selected from the preset criterion (for example, the size of the finally transformed data);
- forming the encrypted communication consisting of two parts, one of which contains the finally transformed data (C_n) being of shorter, equal or longer length than the initial communication, and second one contains accessory data array (F).

A further improvement of a method is characterized by, in each or some cycles, the cycle data (C_i) transformed in the said cycle and (or) accessory data (F_i) for the said cycle are intermixed.

The following improvement of the method is characterized by, in each or some cycles of transformation, the certain part of an accessory data (F_i) for the said cycle is added into the transformed information in the said cycle data. The structural interpretation of stated ideas could be considered on an example of the claimed device.

A device for realizing the method of encrypting data comprises:

- an input unit;
- an output unit, the first input of which is connected to the second output of a commutator, and the second input of which is connected to the output of an accessory data storage;
- a database on the plurality of characteristic functions that transform the initial data with the encoded data, the first input of the said database being connected to the first output of the input unit and the second input - to the output of a random number generator;
- a random number generator, the input of which is connected to the first output of a decision making unit;
- a transformation unit, the first input of which is connected to the second output of the output unit, the second input – to the output of the database, and the third input – to the first output of the commutator;
- a storage for transformed information, the input of which is connected to the first output of the transformation unit;

- a storage for accessory information, the first input of which is connected to the second output of the transformation unit, and the second input – to the second output of the decision making unit;
- a decision making unit, the first input of which is connected to the third output of the input unit, the second input – to the first output of the storage for transformed communication; information;
- a commutator, the first input of which is connected to the second output of the storage for transformed information, and the second input – to the second output of the making decision unit.

Another method of decoding encrypted data comprises the following steps:

- preliminary generating data on plurality of characteristic functions that transform values of all encoded symbols that may be used in the said kind of data with initial symbols, which are identical to the regularities used at encoding;
- extracting, from the encoded communication, of the data (R_i), defining the regularity which is used in the current transformation cycles and connects the values of the encoded communications with the concrete symbols of the transformed data of the current transformation cycle;
- selecting the regularity connecting the values of the encoded communications with the concrete symbols of the transformed data of the current transformation cycle;
- extracting from an accessory data array (F) the accessory data (F_i) for the said transformation cycle;
- transforming the cycle data (C_i) using the selected regularity and the accessory data (F_i) for the said transformation cycle;
- making decision on switching to the next cycle or termination of the transformation;
- the accessory data (F_i) for the said transformation cycle being isolated from the accessory data array (F);
- recovering the cycle data (C_i), which is transformed in the respective cycle, by using the selected regularity and the accessory data (F_i) for the said transformation cycle;
- making decision on switching to the next cycle or termination of the transformation;

- using additionally in each transformation cycle a respective part of the accessory data, as a result of transforming with the use of the selected regularity there is formed the data recovered in the respective cycle.

A further improvement of the method is characterized by:

- in each transformation cycle there is additionally used a respective part of the accessory data and as a result of the transformation with use of the selected regularity there is formed a recovered information in the corresponding cycle communication, the length of which is larger or equal to the length of the communication, resulting from transforming in the previous cycle.

The following improvement of a method is further characterized by the additional use of a respective part of the accessory data in each transformation cycle. As a result of transformation with use of the selected regularity, there is formed recovered information in the respective cycle communication. The length of the recovered information is larger than, equal to or smaller than the length of the communication resulting from transforming in the previous cycle.

One more improvement of the method is characterized in that the respective cycle data (C_i) and/or the accessory data (F_i) for the respective cycle is preliminary unmixed in each cycle or in some cycles.

A device for realizing the method decoding of the communication comprises:

- an input unit 10;
- an output unit 15;
- a database on the plurality of characteristic functions that transform the encoded data with the initial data (2);
- a transformation unit (12);
- a storage for transformed information (14);
- a storage for accessory information (13);
- a making decision unit (11);
- a commutator (8);
- the first input of the storage for accessory information (13) being connected with first output of the input unit (10) and the second input of the accessory data storage

for accessory information (13) being connected with first output a making decision unit (11); the first input of database (2) is connected to the second output of the of the input unit (10), and the second input – to the first output of the storage for accessory information (13); the first input of the storage for transformed information is connected to the third output of the input unit, the second – to the output of the transformation unit; and the third – to the first output of the making decision unit, the first input of the transformation unit is connected to the second output of the storage for accessory information, and the second input to the output of the database, the third to the first output of the storage for transformed information, the second – to the fourth output of the input unit; the first input of the commutator is connected to the second output of the making decision unit, and the second – to the second output of the making decision unit; the output unit is connected to the second commutator output.

With the first output (exit) of the commutator, the first input (entrance) of the decision making unit is connected to the first output (exit) of the storage for transformed information, and second — with the fourth output (exit) of the input unit; the first input (entrance) of the commutator is connected to the second output (exit) of the decision making unit, and second — with the second output (exit) of the storage for transformed information; the output unit is connected to the second output(exit) of the commutator.

A distinctive feature of the new method can be illustrated by the following example. Symbols of the initial alphabet A {al, a2, ..., an} being such, that the binary representation of each symbol has the identical length for the whole alphabet A, are substituted with symbols of the alphabet Bi {bli, b2i, ..., bni} being such, that the binary representation of each symbol may have a various length, the process of such replacement is iterative, i.e. at each i-step for the initial communication there is used a result of the substitution obtained at the i-1step. At each i-step there is used its own substitution alphabet Bi, produced with the help of the function Y_i, selected by a casual mode from a plurality of functions transferred to the addressee beforehand, and at each i-step there is produced the accessory data F_i used for restoring the initial communication. As an additional measure of protecting from cryptanalysis, on each step or on some steps there

may be performed intermixing of the communication resulting from the transformation. In an outcome of such transformation there is produced a finally transformed data (C_n), the length of which may not be less than the length of one symbol of the alphabet Bn, used at the last step of transformation.

Such systems have uncommon properties:

- as a result of transformation of the initial communication there are produced at least two output communications (the finally transformed data (C_n) and the accessory data array (F), each of which separately does not make any sense from the point of view of restoring the initial communication and may be transmitted through a separate data link;
- generally, the length of the transformed communication may have the length of one symbol of the substitution alphabet, for example, if the initial communication has the byte representation, the transformed communication may have the one byte length, regardless of the length and kind of the initial communication;
- at multiple encoding one and the same initial communication, the transformed communication will be various, thereby eliminating a problem of the closed channel for the key data transfer;
- the modification of any symbol in the transformed communication or accessory data brings about the impossibility of restoring the initial communication.

The transformation functions (Y_i) may be preset in the form of a table. For example, in case of representing the initial communication as N-bit binary sequences and transformation of compression of the function (Y_i), can be preset as a set of 2^N triples — $\{(a_k, b_{ik}, f_{ik})\}$, where a_k is an N-bit initial code, b_{ik} is a transformed bit code of a variable length not greater N, and only two values of $\{b_{ik}\}$ have the length of N bit, f_{ik} is the data on the length of the respective b_{ik} in bits. At such representation there exists such submission $(2^N)!(2^N - 1)(2^N - 2)$ of various possible functions of transformation, such that

$\sum_{i=1}^{2^N} L_{ik} = \min L_{ik}$, where L_{ik} is the length of b_{ik} in bits. At N = 8 there is present ≈256!

*254*255 10^{511} of various transformation functions (Y_i). In this case two values of b_{ik} have the one bit length, four values of b_{ik} have the two bit length, eight values of b_{ik} have the

three bit length, sixteen values of b_{ik} have the four bit length, a thirty-two values of b_{ik} have the five bit length, sixty-four values of b_{ik} have the six bit length, one hundred twenty-eight values of b_{ik} have the seven bit length and two values of b_{ik} have the eight bit length.

Then for an arbitrary function Y_i the average length of the transformed communication X will be equal:

$$L(C_i(X,Y_i)) = L(X)\frac{2N+\sum_{n=1}^{N-1} n2^n}{N2^N}$$

and the average length of an accessory data:

$$L(F_i(X,Y_i)) = L(X)\frac{2N+\sum_{n=1}^{N-1} n2^{N-n}}{N2^N}$$

thus, the average compression ratio at one step of transformation will have the values:

$$K_{flags} = \frac{L(F_i(X,Y_i))}{L(X)} = \frac{2N\sum_{n=1}^{N-1} n2^{N-n}}{N2^N}, \textit{ for the transformed communication}$$

$$K_{core} = \frac{L(C_i(X,Y_i))}{L(X)} = \frac{2N+\sum_{n=1}^{N-1} n2^n}{N2^N}, \textit{ for the accessory data.}$$

In particular, for N = 8 we have: K_{core}=777/1024 0.758 K_{flags}=255/1024

At performing transformation M cycles the anticipated average length of the transformed communication will be:

$$L(C(X)) = K_{core}^{M} L(X),$$

and of the accessory data -

$$L(F(X)) = L(X) K_{flags} \sum_{m=0}^{M-1} K_{core}^{m} = K_{flags} L(X) \frac{K_{core}^{m} - 1}{K_{core} - 1}.$$

Accordingly, at performing 10 transformation cycles the average length of the transformed communication at of N = 8 will make approximately 0.067 of the length of the initial communication, and length of the accessory data — 0.97 of the length of the initial communication. The general length will make approximately 1.037 of the initial length, and for 100 transformation cycles — 10^{-12} and 1.04 accordingly.

If at each transformation cycle a S byte of the accessory data is added to the transformed communication, then average length of the transformed communication will be:

$$L(C(X)) = K_{core}^{M} L(X) + S \sum_{m=0}^{M} K_{core}^{M} = K_{core}^{M} L(X) + S \frac{K_{core}^{M+1} - 1}{K_{core} - 1},$$

And length of an accessory data will make:

$$L(F(X)) = \sum_{m=1}^{M} K_{flags} \left(K_{core}^{m} L(F) + S \frac{K_{core}^{m+1} - 1}{K_{core} - 1} \right) = \frac{K_{flags}}{1 - K_{core}} \left(L(X)(1 - K_{core}^{M}) + S \left(M - \frac{K_{core}^{M+2} - 1}{K_{core} - 1} \right) \right)$$

The construction of the claimed device may be realized in various variants realizing the claimed method of encoding data by using the known hardware. All these variants expand technological possibilities of using of the invention.

The main problem of the prototype method is eliminated thereby, i.e. essential increase of the sizes of the encrypted communication in a comparison with the initial one. The disclosed distinctive features of the claimed invention, in a comparison with known engineering solutions, allow designing a device of encoding data providing statistical independence of the encrypted text and the open text, i.e. having properties of the theoretically stable of proof system of cryptography, and not by recurrence of the encrypted communication at repeated encoding of one and the same communication at constant keys.

DETAILED DESCRIPTION

Fig.1 shows a diagram of a device for realizing a claimed method of encoding data. Through an input unit, a database enters pre- generated data on plurality of characteristic functions that transform values of symbols of the initial communication with specific symbols of the encrypted communication for the whole set of symbols of the said kind of the communications. In the course of processing the encrypted data, the input of a making decision unit (3) enters the data on the number (n) of transformation cycles of the initial communication. Before the beginning of the current transformation cycle, the making decision unit (3) transmits a control signal to the random number generator(5), which generates a random number (R_i), transmits it to the database (2) and through the latter to a transformation unit. In accordance with the value of R_i from the database (2), there is selected the transformation function of YR_i which enters the transformation unit (4). The transformation unit (4) calculates the values of (C_i,F_i) - YR_i (X_i, R_i). The value of C_i enters the input of a storage for transformed information (6) from outputs of the transformation unit (4) and the value of F_i enters the input of the storage for accessory information (7). The storage for transformed information (6) transmits a signal on termination of the current cycle of transformation to the making decision unit (3). The making decision unit (3) makes a decision on fulfillment of the next transformation cycle or on terminating the process of transformation. In the case of terminating the process of

transformation, the finally transformed data (C_n) is transmitted through the commutator (8) and the accessory data array ($F = \{F_1, F_2, \ldots, F_n\}$) from the storage for accessory information (7) enters an output unit (9). Otherwise, the cycle data (C_i) through the commutator (8) enters the transformation unit (4) for fulfillment the next cycle of transformation.

Fig. 2 shows the diagram of a device for realizing a claimed method of decoding data. Through the input unit (10) into the database (2) come the previously generated data on plurality of characteristic functions that transform values of symbols of the initial communication with special symbols of the encrypted communication for the whole set of symbols of the said kind of the communications, which are identical to the regularities used at encoding. In the course of restoring the transformed communication through the input unit (10) enters the following data: at the input of a decision making unit (11), - data on the number (n) of transformation cycles of the deencrypted communication; at a storage for accessory information (13) - the accessory data; at a storage for transformed information (14) - the transformed communication. Before the beginning of the current cycle of restoring at the signal of the decision making unit (11) the storage for accessory information (13) yields accessory data (F_i) into a transformation unit (12) and the value of R_i - into the database (2), in accordance with which is selected the function of transformation of YR_i that arrives at the transformation unit (12). The storage for transformed information (14) yields through the commutator (8) cycle data (C_i) into the transformation unit (12). The transformation unit (12) calculates the values of (X_i), YR_i (C_i, F_i). From the output of the transformation unit (12) the restored communication (X_i) arrives into the storage for transformed information (14). At completion of accumulation of the restored communication (X_i), the storage for transformed information (14) sends a signal on termination of the current cycle of restoring into the decision making unit (11). In case of decision-making on the termination of process of transformation, the restored communication (X_i) through the commutator (8) arrives at an output unit (15). Otherwise, from the output of the decision-making unit (11) at the input of the storage for accessory information (13) arrives the signal on yielding of the next portion of the accessory data (F_i, R_i) and the restored communication arrives through the commutator (8) at the transformation unit (12) for fulfillment of the next cycle of restoring.

Bibliographic data of sources of data

1. Victor Gavrish "Practical Guide on Protecting Commercial Secrets". Simferopol, TAVRIDA, 1994, p.35-37.
2. Schmidt M. E., Bransted D.K. "Standard of Data Encoding: Past and Future" Journal of Works of Electronic and Radio Engineers (TIIER), 1988, v.76, no. 5., p. 33-34.
3. GOST 34.11-94 Data Technology, Crypto Graphical Protection of Data, Cash function. M.: Gosstandart of Russia, 1994, 34.11 - 94, p. 3-8.
4. Shannon C.E., "Communication Theory in Secret Systems", Shannon c.E. Works on Data and cybernetics Theory". M.: IL, 1963, p. 333-402, "Theoretically Stable system,", as cited in "An Introduction to Contemporary Cryptology", Proceedings of the IEEE, v. 76. No. 5, May 1998.
5. Vernan. Copher printing telegraph systems for secret wire and radio telegraphic communications.// J Amer. Inst. Elec. Eng., vol. 55, pp. 109-115, 1926.
6. Mischenko V.A, Zakharov V.V. A method of encoding and transfer data and the device for a realization the method // Official Gazette of the Belarusian Patent Office. No.4, part I, 1997
7. Golubev V.V. Computer crimes and protection of data in computing systems // News in life, science and engineering. Part. Computer engineering and use thereof. Protection of data.-M.: Znanie, 1990.